DOI: 10.1021/jm100118x

Journal of Medicinal Chemistry Article

# An Analysis of the Binding Efficiencies of Drugs and Their Leads in Successful Drug Discovery Programs

Emanuele Perola*

*Vertex Pharmaceuticals, 130 Waverly Street, Cambridge, Massachusetts 02139*

*Received January 28, 2010*

In order to investigate the evolution of binding efficiency in successful drug discovery programs, a data set of 60 lead/drug pairs with known binding affinities has been compiled and analyzed. Low-end thresholds for the binding efficiencies of viable leads and drugs have been derived. On average, the drugs in the set are significantly larger and more potent but have similar lipophilicity relative to their originating leads, suggesting that the ability to maintain low levels of lipophilicity while increasing molecular weight is one of the keys to a successful drug discovery program. A number of examples demonstrate that large increases in binding efficiency from leads to more elaborate drugs sharing the same scaffold can be achieved. The importance of dissecting a lead structure to identify the most efficient fragments and the option of sacrificing binding efficiency to optimize other properties are discussed, and relevant examples are highlighted.

## Introduction

Drug discovery is a multidimensional process in which a number of different components must be simultaneously optimized to converge on a viable drug candidate. The process generally begins with the identification of one or more lead molecules that are then optimized through an iterative process of design, modification, and evaluation. Modern drug discovery programs generally rely on the knowledge of a molecular target involved in some critical biological function and on the ability to identify molecules that interact with the target and inhibit its function. The potency measured against the target is often the dominant criterion for lead selection, and it usually remains the primary driver in the early stages of lead optimization. However, modern medicinal chemists have become increasingly cognizant of the importance of modulating the physical properties of their leads early on in the process to avoid being pigeonholed in highly unfavorable regions of property space. One key parameter to be considered in this context is molecular weight, and two conflicting trends have been observed in this regard: (1) a significant increase in molecular weight relative to the initial lead(s) is often required to achieve the necessary level of potency,[1,2] and (2) some of the key properties that determine the druggability of a molecule (e.g., solubility, metabolic stability, oral bioavailability) tend to deteriorate as molecular weight increases beyond a certain point.[3,4] When assessing the viability of a molecule it is therefore important to monitor both potency and molecular weight to ensure that the appropriate balance can be achieved at the end of the optimization process. The concept of binding efficiency provides a convenient means of assessing the relationship between potency and molecular weight,[5] and it is now routinely used as one of the guiding factors in the process of lead selection and in the early stages of lead optimization. Binding efficiency is a measure of the binding energy per unit of mass for a given compound relative to its molecular target. A few different definitions have been proposed for this parameter, which is often referred to as "ligand efficiency" (LE[a]):

$$\mathrm{LE} = \frac{\Delta G_{\mathrm{binding}}}{\text{no. of heavy atoms}}$$

$$\mathrm{LE} = \frac{\mathrm{p}K_{\mathrm{i}},\ \mathrm{p}K_{\mathrm{d}},\ \text{or pIC}_{50}}{\text{no. of heavy atoms}}$$

$$\mathrm{LE} = \frac{\mathrm{p}K_{\mathrm{i}},\ \mathrm{p}K_{\mathrm{d}},\ \text{or pIC}_{50}}{\mathrm{MW\ (kDa)}}$$

Although a rigorous computation of binding efficiency would require the use of a true binding constrant, actual $K_d$ values are rarely measured in drug discovery programs, and $K_i$ or $IC_{50}$ values are commonly used as surrogates. The third definition above, also referred to as "binding efficiency index" (BEI),[6] will be used throughout the rest of this paper. While the importance of maximizing efficiency is now well understood and widely accepted, clear guidelines defining the desirable and the acceptable levels of binding efficiency at various stages of a drug discovery program have yet to be established. The goal of this study was to investigate the ligand efficiency trends in successful drug discovery endeavors, reassess some common assumptions on lead viability and derive new or revised guidelines to be applied in future drug discovery programs. A number of papers have been published on this topic in recent years, and some of them have attempted to analyze the evolution of binding efficiency in the course of lead optimization programs with the goal of establishing some ground rules. A recent study published by researchers at Abbott[7] analyzed the trends observed in a number of internal lead optimization programs and concluded that, once an optimal lead scaffold is selected, the binding efficiency

*Contact information. Phone: (617) 444-6646. Fax: (617) 444-7822. E-mail: emanuele_perola@vrtx.com.

[a] Abbreviations: LE, ligand efficiency; BEI, binding efficiency index; LLE, lipophilic ligand efficiency.

Published on Web 03/17/2010

remains relatively constant during the optimization process if the scaffold is preserved and optimal substitutions are incorporated at each step of the way. On that basis, the minimum size of the optimized molecule can be predicted from the efficiency of the initial lead and the desired affinity. A subsequent study published by researchers at Johnson & Johnson[8] found that the maximum achievable binding efficiency decreases with molecular size. This relationship can be explained with the observations that (a) the relationship between the ligand surface available for interaction and the atom count is not linear, as a proportionally larger number of atoms become partially or totally buried as size and complexity increase and (b) an ideal fit becomes statistically less likely as the molecule becomes more complex.[9,10] If we combine the findings of these two studies, the conclusion is that the binding efficiency of an optimal lead can be maintained at best and will likely decrease during the optimization process, at least when the scaffold is preserved. If that were true, the binding efficiency of a viable lead should be equal to or higher than the efficiency one expects to need in the optimized drug. Verifying the validity of these assumptions and deriving new or revised guidelines were two of the main goals of this study. The design of the study entailed the following steps:

(1) generation of a database of lead/drug pairs with known binding affinities;
(2) calculation of binding efficiencies and other relevant descriptors;
(3) analysis of the variations of binding efficiency from beginning to end of the drug discovery process and their dependency on other parameters;
(4) reassessment of published guidelines/dogma on lead viability;
(5) establishment of new/updated guidelines for lead selection/optimization.

## Methods

A database of lead/drug pairs was generated as a result of a thorough search of the literature, online sources detailing names and structures of approved drugs, and existing drug databases. The lead/drug pairs identified in the search were incorporated in the database if the following criteria were satisfied:

(1) The lead was reported or clearly identifiable as compound no. 1 in the discovery path.
(2) The binding affinity (as $K_i$, $K_d$, or $IC_{50}$) was reported for both lead and drug.
(3) The same assay was used to measure the affinity of lead and drug.
(4) The lead was not an approved drug at the time of discovery.
(5) If multiple drugs were based on the same lead, only one pair was included.
(6) Withdrawn drugs were excluded.

The search resulted in the identification of 60 lead/drug pairs satisfying the above criteria. Four of the drugs (benazeprilat, fosinoprilat, dabigatran and oseltamivir carboxylate) are administered as prodrugs. Two of the 60 drugs were approved between 1978 and 1990, while the remaining 58 were approved between 1991 and 2008. The discoveries took place in 40 different companies, and the targets encompass 23 enzymes and 16 receptors.

The calculations of ClogP and $\mathrm{ClogD}_{7.4}$ have been performed with the calculator plugins from ChemAxon.[11] The computation of the maximum common substructure between leads and corresponding drugs has been performed using an internally developed program. Property distributions, histograms, and plots have been generated with Microsoft Excel. The snapshots of the 3D structures have been generated with PyMOL.[12]

**Table 1.** Breakdown of Lead Identification Methods for the 60 Leads in the Data Set

| Source | No. of leads |
|---|---|
| literature compound | 15 |
| HTS | 14 |
| scaffold morphing from literature or competitor compound | 11 |
| substrate or transition state analog | 10 |
| diversity screen | 5 |
| pharmacophore screen | 3 |
| screen against related enzyme | 1 |
| derivative of literature compound | 1 |

## Results and Discussion

**Quality and Scope of the Data Set.** Large amounts of data are available today for the vast majority of marketed drugs from databases, review articles and a variety of other sources. However, compiling a sizable data set of lead/drug pairs with the corresponding binding affinities proved to be challenging for a number of reasons. First, drug discovery programs are often poorly documented in the literature, and the published reports often lack clear and detailed information about the early stages of discovery. Second, prior to the past 3 decades, much of drug discovery was not target-driven, and binding data was rarely used as the guiding factor. Third, many drugs target complex systems for which binding data is difficult to obtain. Fourth, the assays often evolve in the course of a program and as a result the assays used to test the initial lead and the final drug can be different. And fifth, a surprisingly high number of drugs are routinely approved when the mechanism of action and/or the exact molecular target are still unknown (at least 15 examples can be found in the 2007−2009 period alone[13]). For these reasons the data set compiled for this study could only cover a small fraction of the currently available drugs. However, the vast majority of the drugs in the data set were approved in the past 2 decades. Considering that the number of novel small molecule drugs approved in the same time period can be estimated to be around 250−300, the data set used here can be regarded as a highly representative sample of modern drug discovery programs. The drugs in the data set are structurally diverse and encompass a large number of targets and therapeutic areas. Additionally, the criteria applied in the selection of viable drug/lead pairs minimized the bias toward particular compound classes that were developed from common leads and maintained a clear separation between leads and drugs by excluding cases where a drug was evolved from another drug.

**Lead Identification Methods.** The breakdown of the identification methods for the 60 leads analyzed in this study is reported in Table 1. It is remarkable that one-quarter of the leads were compounds previously reported in the literature, while high throughput screening follows closely as the second most common source of drug leads. Eighteen percent of the leads were designed on the basis of targeted modifications (morphing) of the core scaffolds of compounds from the literature, patents or existing drugs. The difference between this category and that of literature leads is that in the former the core scaffold of the source was deliberately modified from the start resulting in a different chemical class, often with initial loss of activity but gain of intellectual

**Table 2.** Number of Drugs in the Database Discovered with the Aid of Structure-Based Design: Breakdown by Target

| Target | No. of drugs |
|---|---|
| HIV-1 protease | 4 |
| influenza neuraminidase | 2 |
| renin | 1 |
| DPP-IV | 1 |
| HIV-1 reverse transcriptase | 1 |
| VEGFR-2 kinase | 1 |
| LCK | 1 |

property, while in the latter the original scaffold was retained for at least part of the optimization process. The fourth main category of leads is that of substrate or transition state analogs, which accounts for 16.7% of the current set. Other lead identification methods were significantly less common. Overall, over 60% of the leads were based on molecules that were previously known to be related to the target by virtue of being inhibitors or substrates, while less than 40% were completely novel molecules identified by screening approaches. In this group, only 8 lead compounds were identified through the use of computational methods, and 5 of those 8 emerged from diversity screens. No lead in this set was identified through a target-based virtual screen, suggesting that for some reason docking-based virtual screening methodologies are not yet making the desired impact on lead discovery.

**Impact of Structure-Based Design.** The advances in the field of protein crystallography have made the resolution of the three-dimensional structures of protein targets a routine task for soluble proteins and a realistic albeit challenging possibility for membrane-bound targets. As a consequence, the incorporation of structure-based design approaches in the drug discovery cycle has taken place to different degrees in the majority of pharmaceutical and biotech companies. A search of the Protein Data Bank reveals that the 3D structure of the target is now known for 36 of the drugs in the database generated in this study. However, a thorough review of the articles describing the path to the discovery of these drugs shows that structure-based design methods were only applied in 11 cases. The reason is that in the remaining 25 cases the structure of the target was solved after the end of the program or when the program was advanced enough not to require structural input for the final lead optimization steps. The targets to which structure-based methods were successfully applied are listed in Table 2. HIV-1 protease remains the prototypical target for structure-based drug design, with the largest number of successful applications reported to date. However, only about half of the HIV-1 protease inhibitor drugs in the database were discovered with the help of structure-based approaches, while the earliest to reach approval were the result of more traditional drug discovery approaches. This partly dispels the notion that structure-based design was responsible for the HIV treatment breakthroughs since the beginning. Another misconception that is dispelled by this analysis is that kinase inhibitor drugs generally resulted from structure-guided design programs. The database in this case contains six of the eight currently approved kinase inhibitor drugs, and based on the related literature structure-based methods were only applied in the discovery of two of them (dasatinib[14] and lapatinib[15]). For example, nowhere in the original reports is there a mention of target structure in the discovery of imatinib,[16−18] which is often referred to as one of the recent successes of structure-based drug design. Notably renin, a target for which a huge amount of structural and modeling work has been reported, has finally born fruit in terms of producing an approved drug, thus justifying the emphasis given to those studies and the importance of structural information for this challenging protease target.

**Molecular Weight and Affinity Distributions.** The molecular weight distributions for the leads and the drugs in the data set analyzed in this study are reported in Figure 1A. The distribution for the drugs is wider and fairly even between 200 and 600 Da, while three-quarters of the leads fall within the narrower range of 200−400. The median values are 328 for the leads and 436 for the drugs. The histogram in Figure 1B illustrates the distribution of the molecular weight differences within drug/lead pairs. The drug is larger than the lead in 82% of the cases, and the average molecular weight difference is 89.5 Da, which is slightly smaller than the difference of the medians when drugs and leads are analyzed as two separate groups. Figure 2A illustrates the distribution of the binding affinities for the two groups. In this case the distribution is wider for the leads than it is for the drugs, as the leads populate the low affinity region which is not an option for the drugs. The distribution for the drugs is shifted toward lower $K_i$ values: the majority of the drugs have $K_i$ values between 0.1 and 100 nM, while the leads are centered between 10 nM and 10 $\mu$M. The adjacent graph (Figure 2B) shows the distribution of the affinity differences within drug/lead pairs, expressed as $\Delta pK_i$. The drug is more potent than the corresponding lead in 90% of the cases, and the average difference in potency is 2 log units (100-fold), which is almost identical to the difference of the median $pK_i$ values of drugs and leads as separate groups. In summary, drugs are on average about 100 Da larger and 100 times more potent than the corresponding leads, and in the vast majority of the cases the drug is indeed both larger and more potent than its originating lead. Exceptions are uncommon but they are definitely observed.

**Ligand Efficiency Distributions.** The distribution of the ligand efficiencies for leads and drugs is illustrated in Figure 3A. The distribution is wider for the leads, as they populate the lowest end of the efficiency spectrum which is not viable for the drugs. Leads with ligand efficiencies as low as 6.8 (which correspond to a MW of 631 and a $K_i$ of 53 $\mu$M) have been successfully optimized to approved drugs, while no drug in this set has an efficiency lower than 10. Ninety percent of the leads have efficiencies above 12.4, while 90% of the drugs have efficiencies above 14.7. These values could be used as thresholds for the minimal desired efficiencies when defining target profiles for a starting point and an end point in a drug discovery program. Interestingly, the median values for leads and drugs are identical at 18.4. Although this observation alone would appear to be consistent with the conservation of binding efficiency observed in the Abbott study for idealized lead optimization programs, the bar graph in Figure 3B shows how variable the ratio of the ligand efficiencies can be within corresponding drug/lead pairs. Efficiency changes of 20% or more in either direction are observed in almost half of the pairs, thus showing that large increases or decreases in ligand efficiency are common throughout lead optimization programs. On average, drugs are more efficient than their originating leads by 11%, thus showing that the pairwise analysis in this case provides slightly different results than the analysis of leads and drugs as separate groups. In this data set the binding efficiency of

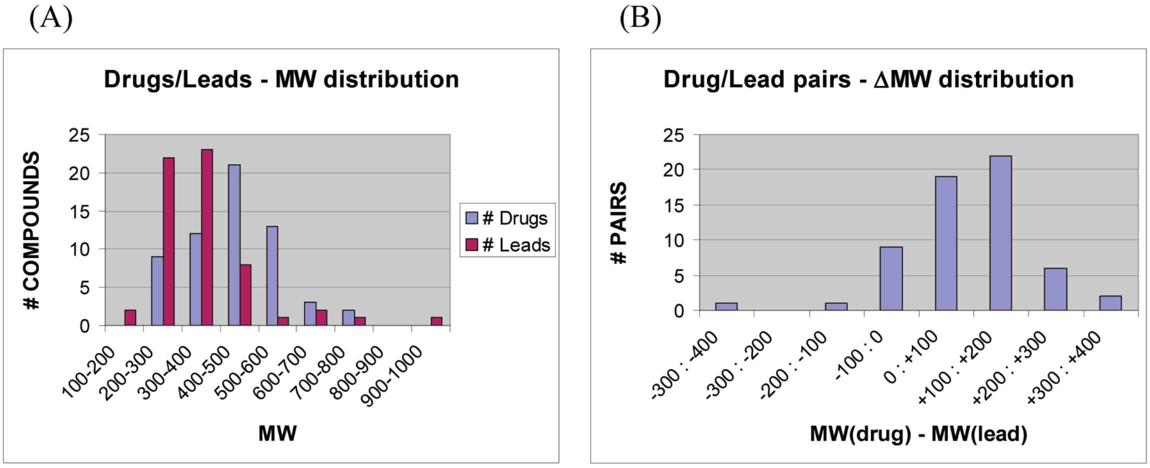


**Figure 1.** (A) Molecular weight distribution for drugs and leads as separate groups. (B) Distribution of the molecular weight differences between drugs and corresponding leads.

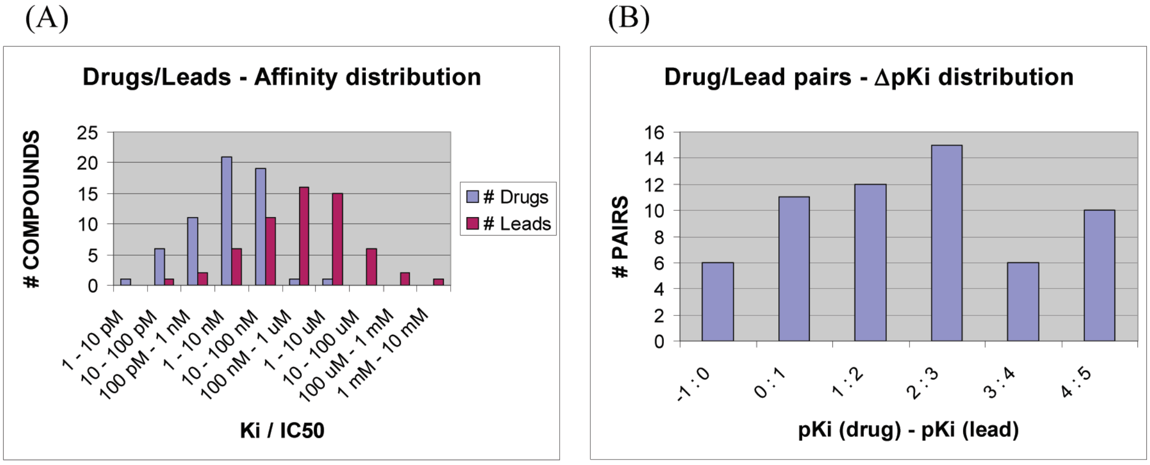


**Figure 2.** (A) Distribution of binding affinities for drugs and leads as separate groups. $K_i$, $K_d$, or $IC_{50}$ values are used depending on the data reported in the original papers. (B) Distribution of the affinity differences between drugs and corresponding leads, expressed as $\Delta pK_i$, $\Delta pK_d$, or $\Delta IC_{50}$.

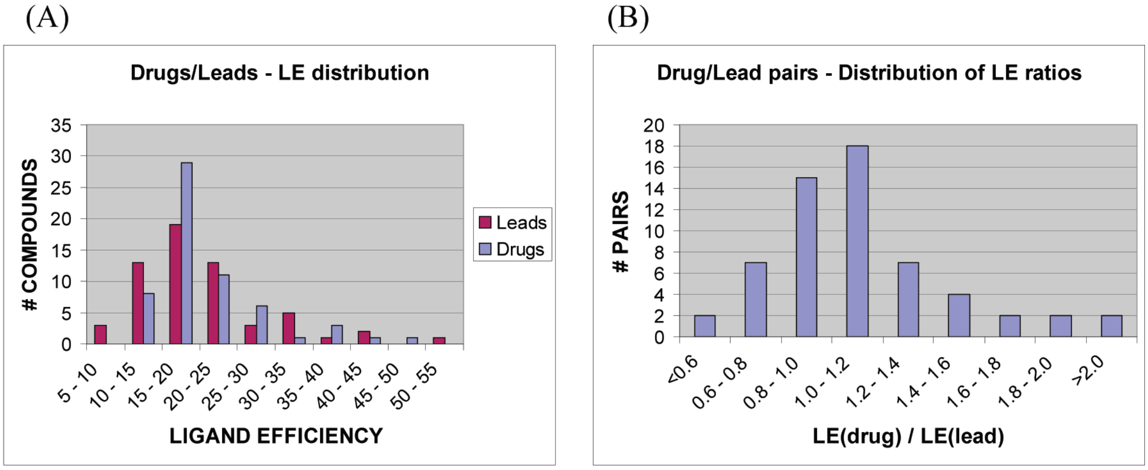


**Figure 3.** (A) Distribution of binding efficiencies for drugs and leads as separate groups. (B) Distribution of the binding efficiency ratios between drugs and corresponding leads.

the drug is higher than the binding efficiency of the corresponding lead in 58% of the cases. Figure 4 illustrates the distribution of the ligand efficiencies for the leads divided by identification method. Leads derived from the literature have the widest distribution and include the most efficient, while substrate and transition state analogs tend to have the lowest efficiencies. A plausible explanation for the latter is that substrate and transition state analogs are often the only viable option for targets like proteases with large and predominantly solvent-exposed active sites. In these cases a

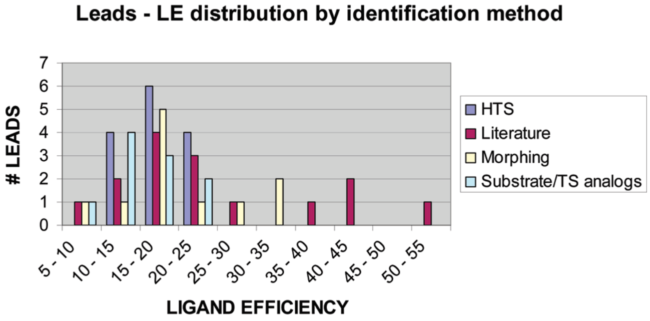


**Figure 4.** Distribution of the binding efficiencies for the leads in the data set broken down by identification method. Only the four most prevalent lead identification approaches are included.

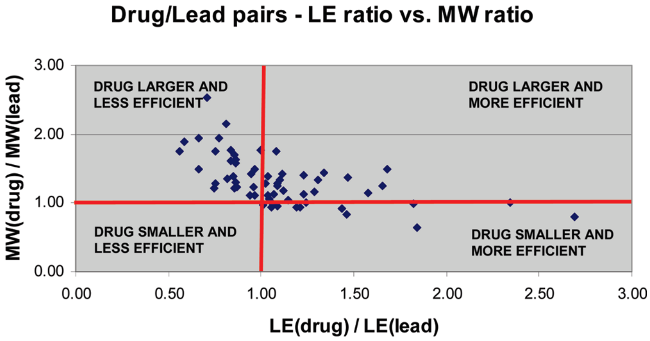


**Figure 5.** Plot of binding efficiency ratio vs molecular weight ratio for the 60 drug/lead pairs in the data set.

large contact surface is necessary to build up significant amounts of binding energy due to the lack of enclosure, thus requiring large ligands with inevitably low binding efficiencies. Leads identified by high-throughput screening have the narrowest distribution of ligand efficiency, every example in the set having efficiency between 10 and 25.

As mentioned in the Introduction, a study published by Johnson & Johnson indicated that the maximum achievable binding efficiency decreases as the size of the molecule increases. The plot in Figure 5 illustrates the correlation between molecular weight ratio and efficiency ratio within drug/lead pairs. The trend is consistent with the Johnson & Johnson findings, as increases in molecular weight tend to result in decreases in efficiency. However, a closer inspection of the plot highlights a few additional points: (1) All the pairs for which the efficiency ratio is lower than 1 (drug less efficient than corresponding lead) are situated in the upper left quadrant, where the molecular weight ratio is higher than 1 (drug larger than corresponding lead). In other words, a decrease in efficiency going from lead to drug always occurs when the molecular weight increases. (2) The opposite is not true. Efficiency increases going from lead to drug can and do occur when the molecular weight increases as well, as evidenced by the pairs that populate the upper right quadrant (drug both larger and more efficient than corresponding lead). In fact, 69% of the drugs that are more efficient than the corresponding lead are also larger, thus showing that an increase in size does not inevitably lead to a decrease in efficiency, and that in ideal lead optimization paths (which we can assume are approximated by successful lead optimization endeavors) the efficiency can indeed be increased, in some cases very significantly.

**Variations in Lipophilicity: Lipophilic Ligand Efficiency.** Another common belief is that the lipophilicity tends to increase in the course lead optimization programs, and as a result drugs tend to have higher ClogP/ClogD relative to the corresponding leads. The histograms in Figure 6 tell a different story. The distributions of ClogP for leads and drugs as separate groups are almost identical, and the median values are 3.14 and 3.04, respectively. Over 75% of leads and drugs have ClogP between 0 and 6. The variation of ClogP within individual drug/lead pairs is within 2 units in the vast majority of the cases, and the median difference is 0, consistent with the difference of the medians for the two groups. The distributions of ClogD at pH 7.4 for leads and drugs as separate groups are also very similar, and the median values are 2.04 and 2.11, respectively (data not shown). The variations of $\mathrm{ClogD}_{7.4}$ within individual drug/lead pairs are also within 2 log units in the vast majority of the cases, and the median difference is 0.09. We can therefore conclude that the average lipophilicity of drugs and corresponding leads is virtually identical. Considering that drugs tend to have a significantly higher molecular weight than the corresponding leads, it is clear that the added molecular weight must carry an adequate proportion of polar and lipophilic groups.

A variation of the concept of ligand efficiency has been recently introduced to measure the extent to which binding is driven by specific interactions between ligand and protein as opposed to simple hydrophobic effects (binding in a protein cavity as a way to escape from solvent).[19] This new metric is defined as "lipophilic ligand efficiency" (LLE), and it is simply expressed as the difference between $\mathrm{p}K_\mathrm{i}$ (or $\mathrm{pIC}_{50}$) and ClogP (or $\mathrm{ClogD}_{7.4}$):

$$\mathrm{LLE} = \mathrm{p}K_\mathrm{i}\ (\text{or } \mathrm{pIC}_{50}) - \mathrm{ClogP}\ (\text{or } \mathrm{ClogD}_{7.4})$$

This metric can be interpreted as a measure of how effectively "grease" is utilized to achieve affinity (less grease, more potency → higher LLE), or alternatively, it can be viewed as a measure of how effectively lipophilicity was minimized in the process of optimizing potency. It could be argued that the binding affinity of a molecule for its target could be broken down into a nonspecific component, which is the tendency to transfer from water to a more lipophilic environment, and a specific component, which is the tendency to bind to a particular protein site as a result of specific interactions. It can be expected that a larger specific component may lead to a more selective compound, and that would be consistent with the findings from a recent study, where the promiscuity of a set of molecules (and therefore the likelihood of undesired activities) was shown to be directly proportional to ClogP (the nonspecific component).[19] The distributions of lipophilic efficiencies for leads and drugs based on ClogP are illustrated in Figure 7. The distribution is shifted toward higher values for the drugs relative to the leads whether ClogP or $\mathrm{ClogD}_{7.4}$ is used in the calculation. The difference between the median lipophilic efficiencies of drugs and leads as separate groups is 2.00 and 2.18 log units depending on whether ClogP or $\mathrm{ClogD}_{7.4}$ is used in the calculation. When the individual drug/lead pairs are analyzed, the drug has a higher LLE in 80% of the cases according to both metrics, and the medians of the pairwise differences are 1.53 and 1.56 log units when ClogP and $\mathrm{ClogD}_{7.4}$ are used, respectively. This trend reflects the fact that drugs are generally more potent than their corresponding leads while having similar lipophilicity. This observation can be translated into this important lesson: one of the keys to a successful lead optimization program is the ability to

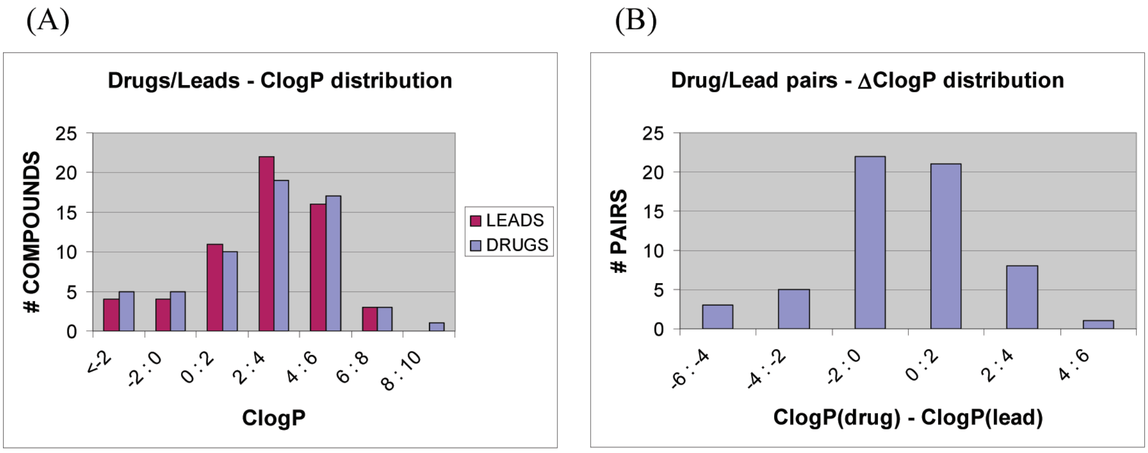


**Figure 6.** (A) Distribution of ClogP for drugs and leads as separate groups. (B) Distribution of the ClogP differences between drugs and corresponding leads.

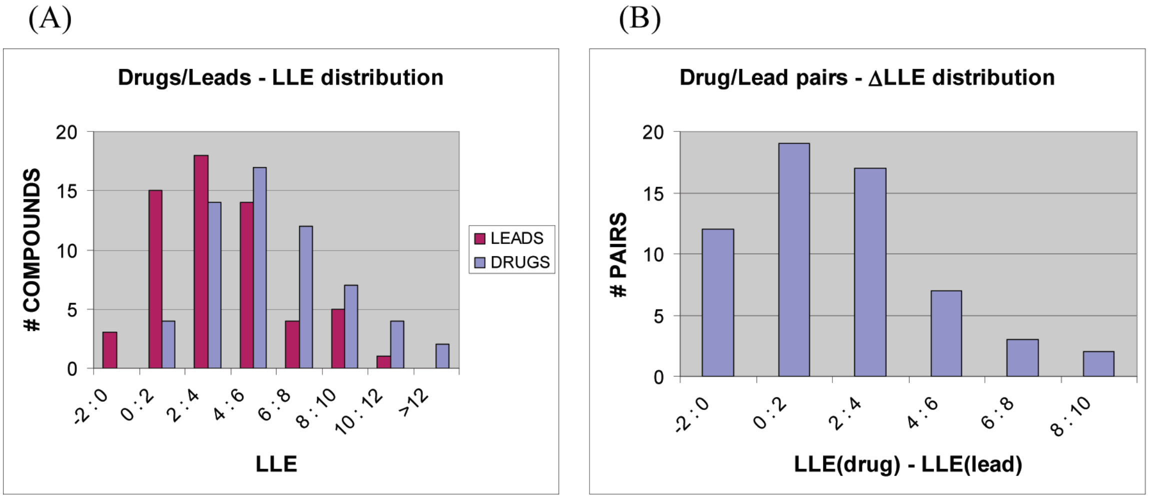


**Figure 7.** (A) Distribution of lipophilic ligand efficiencies for drugs and leads as separate groups, expressed as a function of ClogP. (B) Distribution of the lipophilic ligand efficiency differences between drugs and corresponding leads, expressed as a function of ClogP.

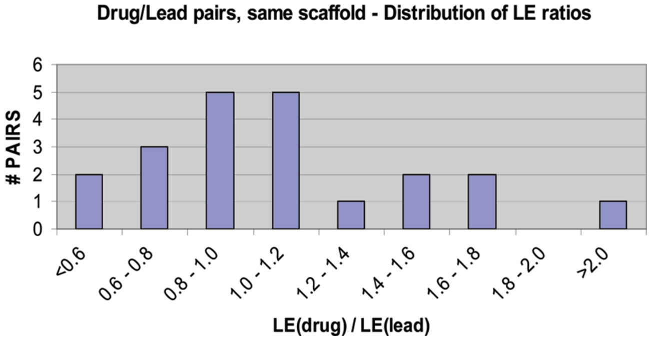
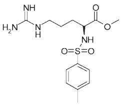
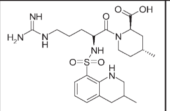
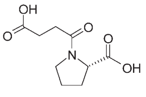
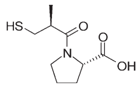
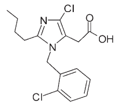
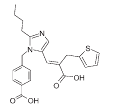
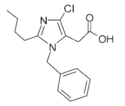
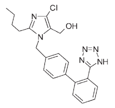
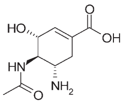
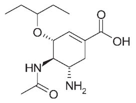
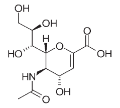
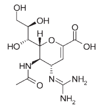


**Figure 8.** Distribution of the binding efficiency ratios between drugs and corresponding leads in pairs where the entire framework of the lead is conserved in the drug.

maintain a relatively low level of lipophilicity in spite of the inevitable increase in molecular weight that is often required to achieve the necessary level of potency.

**Scaffold Conservation and Impact on Ligand Efficiency.** The trends described so far are totally independent of the degree of structural variation occurring between the leads and the corresponding drugs. Arguably, the variations in binding efficiency from the beginning to the end of the lead optimization process are much more relevant and interpretable when the scaffold is preserved throughout the process. The conservation of binding efficiency highlighted by Hajduk and colleagues in internal Abbott programs was based on series where the core of the molecule remained constant. In the data set analyzed in the present study the conservation of the molecular framework, defined as the totality of the rings combined with the minimal set of linkers necessary to connect them,[20] was analyzed. The molecular framework of the lead was fully preserved in the final drug in 21 out of 60 pairs (35%). Ten of the 21 drugs in question preserve the entire structure of the lead when the heavy atoms are considered, and 31 drugs out of the 60 preserve at least 80% of the structure of the lead. The distribution of the ligand efficiency ratios for the 21 pairs in which the scaffold is preserved is illustrated in Figure 8. Notably, there is a high percentage of pairs in which the efficiency varies by 20% or more in either direction from lead to drug. More significantly, 11 of the 21 drugs have higher efficiencies relative to the corresponding leads in spite of the fact that 10 out of 11 also have higher molecular weight. Even more significantly, in 6 of these pairs the ligand efficiency of the drug exceeds the efficiency of the lead by 30% or more, in stark contrast with the conclusions from the Abbott study, according to which the efficiency of

**Table 3.** Drug/Lead Pairs in Which the Core Scaffold of the Lead Is Preserved and LE(Drug) > 1.3 × LE(Lead): Structures and Binding Data

| Lead Structure | Drug Structure | Drug Name | Target | LE (lead) | LE (drug) | LE ratio | Ki ($\mu$M)[a] (lead) | Ki ($\mu$M)[a] (drug) |
|---|---|---|---|---|---|---|---|---|
|  |  | Argatroban | Thrombin | 8.8 | 14.7 | 1.68 | 1000 | 0.032 |
|  |  | Captopril | Angiotensin Converting Enzyme | 15.0 | 35.2 | 2.34 | 590 | 0.023 |
|  |  | Eprosartan | Angiotensin II receptor | 12.8 | 21.2 | 1.66 | 43 | 0.001 |
|  |  | Losartan | Angiotensin II receptor | 12.5 | 18.3 | 1.47 | 150 | 0.019 |
|  |  | Oseltamivir | Influenza neuraminidase | 24.3 | 31.7 | 1.30 | 6.3 | 0.001 |
|  |  | Zanamivir | Influenza neuraminidase | 18.5 | 29.2 | 1.58 | 4 | 0.0002 |

[a] $K_i$, $IC_{50}$, or $K_d$ values are reported depending on availability. The same type of binding data is reported for each individual drug/lead pair.

an optimal scaffold can at best be retained when the ideal optimization path is taken. These results show that at least in some cases a lead based on an optimal scaffold can be evolved into a drug that is significantly more efficient while retaining the same scaffold and increasing the size. One could argue that it cannot be unequivocally proved that the six leads were based on the optimal scaffold for the corresponding binding site. On the other hand, if the scaffold is still present in the final approved drug, that is the closest one could get to proving that point. Structures, targets and binding data for the six pairs in question are reported in Table 3. The targets are three enzymes (influenza neuraminidase, ACE and thrombin) and one receptor (angiotensin II receptor). Three of the drugs have similar size and overall structure as the corresponding lead, while the other three drugs are significantly larger and less similar relative to their leads. In three cases the drug has one additional charged group relative to the lead, in two cases the formal charge is the same and in one case the lead has one additional charged group. Overall the analysis of the 2D structures and the targets does not highlight any features that are common to all six pairs, although charge emerges as a possible factor. The crystal structures of 4 of the 6 drugs in complex with their target are available from the PDB and provide an opportunity to analyze the impact of the structural difference between lead and drug on the interaction with the target and to dissect the factors responsible for the boost in efficiency. Figure 9A illustrates a snapshot of the complex between argatroban and its target thrombin. The drug contains two methylated piperidine rings that are absent in the lead, and one of them is carboxylated. Assuming that the common portions of lead and drug maintain the same orientation, position and interactions, the additional fragments present in the drug engage in two hydrogen bonds (one involving a negatively charged acceptor) and a number of hydrophobic contacts. Each of these interactions may have contributed to the affinity jump, but no specific interaction clearly stands out as the main contributor. The complex between captopril and its target ACE is illustrated in Figure 9B. In this case both lead and drug are very small molecules and the only structural differences between the two are a change of the zinc binding group from carboxylate to thiol and the addition of a methyl group. The affinity skyrockets going from 590 $\mu$M to 23 nM, and due to the similarity in size binding efficiency increases proportionally. The striking impact of these small changes on potency is likely due to the change of the zinc binding moiety. It is not uncommon to see dramatic changes in affinity upon variation of a metal-binding warhead, and that is what happens in this case. The thiol group may achieve a better placement relative to the zinc ion as well as a better interaction network around the metal. The reduced desolvation energy required for binding of a thiol (predominantly neutral in solution) relative to a carboxylic acid (predominantly charged) may play a role as well. The added methyl group makes additional hydrophobic contacts that may further contribute to the increase in affinity. The complex between oseltamivir carboxylate and influenza neuraminidase is depicted in Figure 9C. In this case the drug retains the entire structure of the lead, the only difference being the alkylation of a hydroxyl group with a 3-pentyl group. The etherification leads to a 6000-fold increase in potency and a 30% increase in binding efficiency. The interpretation of the causes is not obvious. Although the pentyl group makes a number of additional van der Waals contacts with the protein, the pentyl-binding region is mainly polar and solvent exposed, and one would not expect such a dramatic effect on potency simply as a result of hydrophobic interactions. The reduced desolvation energy resulting from

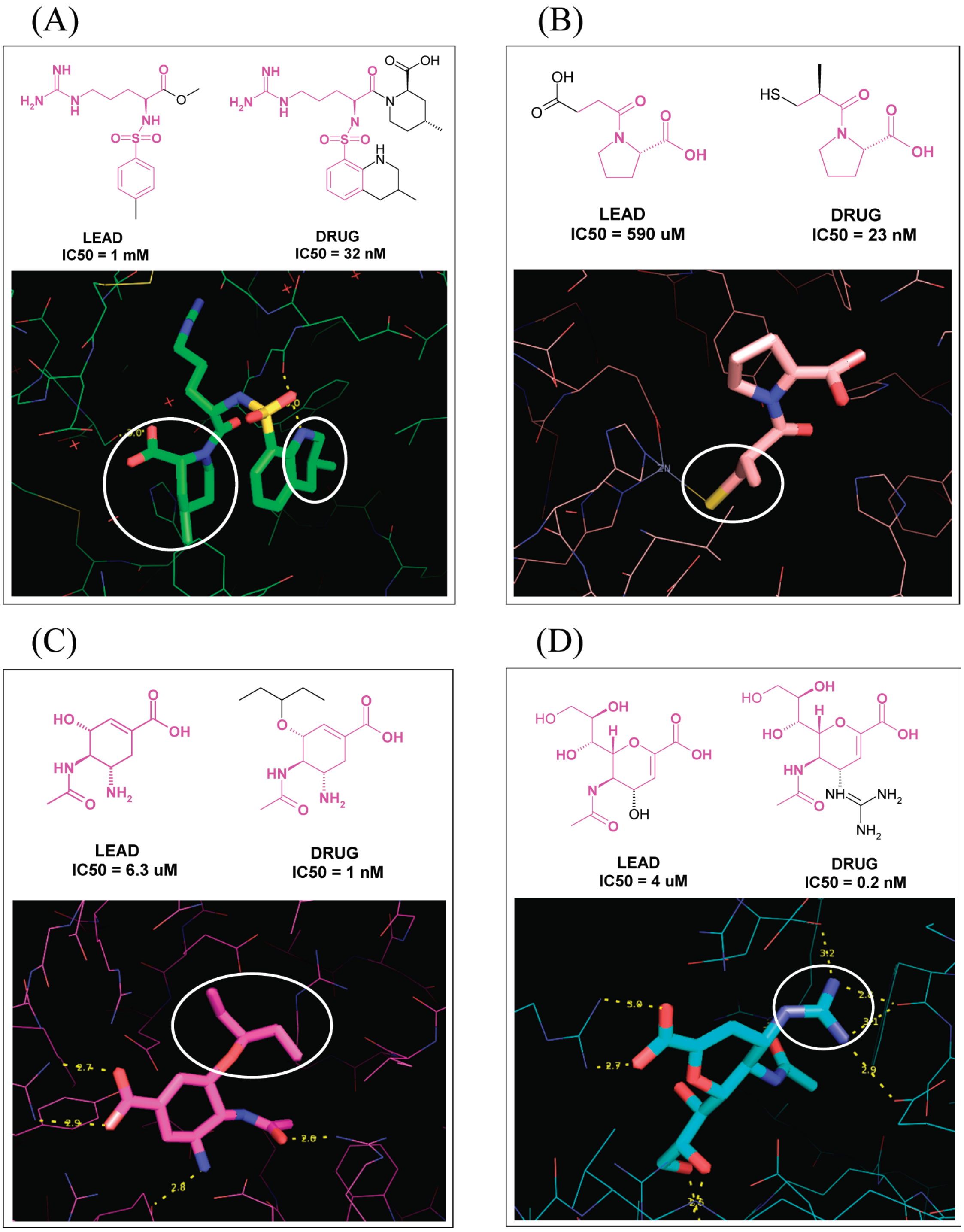
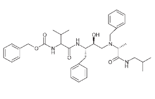
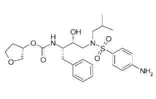
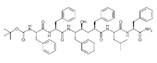
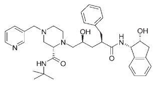
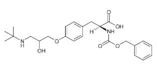
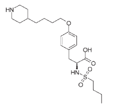
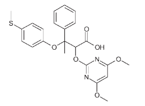
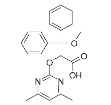
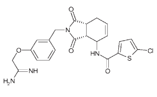
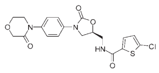
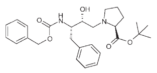
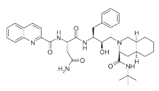


**Figure 9.** Illustration of the four drug/lead pairs with LE(drug)/LE(lead) > 1.3 for which the crystal structure of the drug/target complex is available. The four drugs are argatroban (A), captopril (B), oseltamivir (C) and zanamivir (D). Each quadrant is organized as follows. Top: 2D structures of drug and originating lead with the corresponding binding affinities. The common substructure is colored in magenta. Bottom: snapshot of the crystal structure of the drug in complex with its target. The parts of the drug that are not present in the originating lead are contained within the white circles.

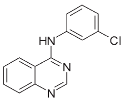
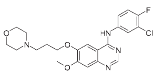
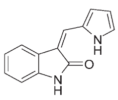
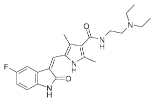
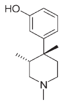
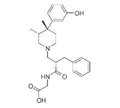
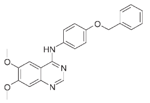
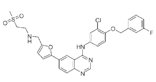
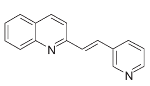
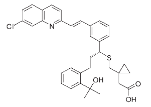
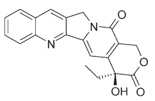
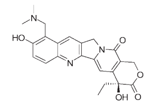
the alkylation of the hydroxyl group may play a role, and there is certainly a chance that high energy water molecules may have been displaced by the pentyl group. Unfortunately the structure of the lead in complex with neuraminidase is not publicly available and possible effects due to conformational changes in the protein active site cannot be fully assessed. The complex between the same target and the related drug zanamivir is depicted in Figure 9D. In this case the difference between drug and lead is in the replacement of a hydroxyl group with a guanidine, which results in a 20000-fold increase in potency and a 57% increase in binding efficiency. The explanation for the potency boost is more apparent for this pair: the hydroxyl group in the lead molecule only makes a weak hydrogen bond to a glutamate side chain, while the guanidine makes four strong hydrogen bonds to the protein, one of which involves two charged partners. These hydrogen bonds take place in a relatively enclosed region, which contributes to their effectiveness.

**Table 4.** Drug/Lead Pairs in Which the Core Scaffold of the Lead Is Not Preserved and LE(Drug) > 1.3 × LE(Lead): Structures and Binding Data

| Lead Structure | Drug Structure | Drug Name | Target | LE (lead) | LE (drug) | LE ratio | Ki (μM)[a] (lead) | Ki (μM)[a] (drug) |
|---|---|---|---|---|---|---|---|---|
|  |  | Amprenavir | HIV-1 protease | 6.8 | 18.2 | 2.69 | 53 | 0.0006 |
|  |  | Indinavir | HIV-1 protease | 9.3 | 17.1 | 1.84 | 0.001 | 0.00003 |
|  |  | Tirofiban | Fibrinogen receptor | 10.4 | 18.8 | 1.82 | 25 | 0.005 |
|  |  | Ambrisentan | Endothelin-A receptor | 14.9 | 21.7 | 1.46 | 0.16 | 0.006 |
|  |  | Rivaroxaban | Factor Xa | 14.6 | 21.0 | 1.44 | 0.12 | 0.0007 |
|  |  | Saquinavir | HIV-1 protease | 11.1 | 14.8 | 1.34 | 6.5 | 0.00012 |

[a] $K_i$, $IC_{50}$, or $K_d$ values are reported depending on availability. The same type of binding data is reported for each individual drug/lead pair.

**Table 5.** Drug/Lead Pairs in Which the Core Scaffold of the Lead Is Preserved and LE(Lead) > 1.3 × LE(Drug): Structures and Binding Data

| Lead Structure | Drug Structure | Drug Name | Target | LE (lead) | LE (drug) | LE ratio | Ki (μM)[a] (lead) | Ki (μM)[a] (drug) |
|---|---|---|---|---|---|---|---|---|
|  |  | Gefitinib | EGFR tyrosine kinase | 30.5 | 17.1 | 0.56 | 0.016 | 0.023 |
|  |  | Sunitinib | VEGF-R2 kinase | 30.5 | 17.8 | 0.58 | 0.39 | 0.08 |
|  |  | Alvimopan | u opioid receptor | 32.4 | 21.5 | 0.66 | 0.08 | 0.00077 |
|  |  | Lapatinib | EGFR kinase | 19.9 | 13.2 | 0.66 | 0.02 | 0.022 |
|  |  | Montelukast | Leukotriene D4 receptor | 22.5 | 15.9 | 0.71 | 6 | 0.0005 |
|  |  | Topotecan | Topoisomerase I | 17.5 | 13.0 | 0.75 | 0.82 | 3.2 |

[a] $K_i$, $IC_{50}$, or $K_d$ values are reported depending on availability. The same type of binding data is reported for each individual drug/lead pair.

The target-bound structure is available for the lead as well and shows that the active site conformation and the orientation/position of the scaffold are fully preserved between the two complexes, thus confirming the validity of the interpretation. Notably, structure-based design approaches were used in the discovery of 2 of the 6 drugs described above (zanamivir and oseltamivir carboxylate), as documented by the corresponding accounts.[21,22] Overall this analysis does not reveal a common overarching pattern shared in all the pairs where a large efficiency boost was achieved, but it highlights some contributing factors. First, the knowledge of the structure of the target can greatly help to identify the hot

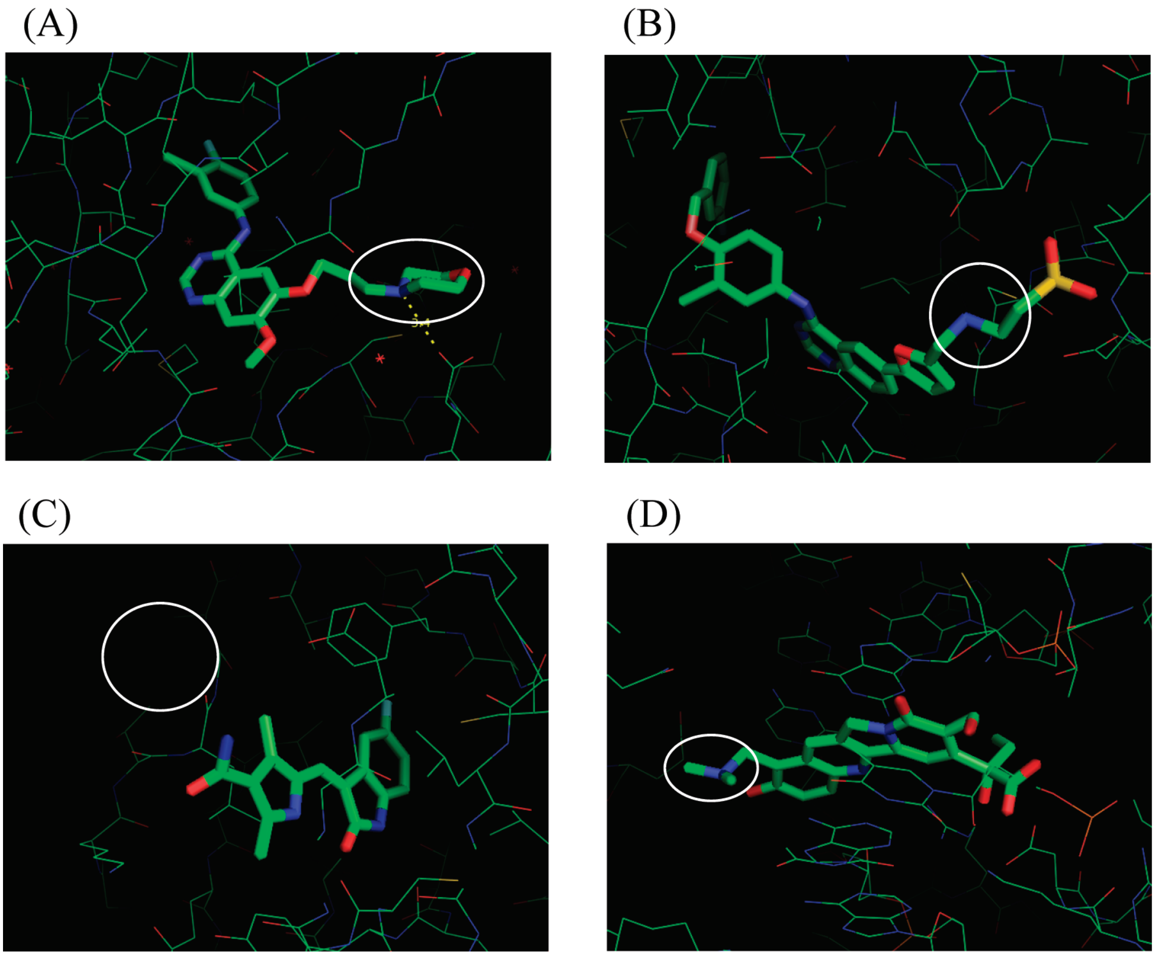


**Figure 10.** Snapshots from the crystal structures of four complexes: gefitinib/EGFR kinase (A), lapatinib/EGFR kinase (B), sunitinib/VEGFR2 kinase (C) and topotecan/topoisomerase I (D). The white circles highlight the positions of the basic amino groups that are present in the four drugs but not in the corresponding leads. In the case of sunitinib, the portion of the molecule containing the amino group was not visible in the crystal structure, and the white circle indicates its approximate location.

spots in the active site and fill them with potentially high-impact moieties. Another recurring theme is the presence in the drugs of charged groups, which under the appropriate set of conditions can make highly effective interactions. Other types of interactions and components can also be responsible for massive increases in efficiency, and additional studies may reveal that displacement of high energy water molecules may be a key factor in some of the least interpretable cases.

Further analysis of the complete data set shows that there are 6 additional drug/lead pairs where the scaffold is not conserved and the binding efficiency also increases by 30% or more. Structures, targets and binding data for these 6 pairs are reported in Table 4. Interestingly, 3 of these 6 cases involve HIV protease inhibitors. In 5 of the 6 cases the drug is smaller than the corresponding lead molecule, thus showing that at least in part the boost in efficiency was the result of removal of ineffective or unnecessary fragments from the lead structure. The HIV protease inhibitor saquinavir is the only drug in this subset that is larger than the corresponding lead, and in this case there is a significant scaffold change and the similarity between lead and drug is relatively low. In the other 5 cases the efficiency boost can be attributed to a cleanup of the lead structure, reengineering of the scaffold and other significant structural changes. This subset highlights the importance of dissecting a lead structure to identify the most efficient fragment(s) and remove the unnecessary atoms in order to begin the optimization process from the most favorable starting point possible.

It is also informative to analyze the cases where the optimization process led to a large decrease in binding efficiency. The complete data set contains 6 drug/lead pairs in which the scaffold is conserved and the efficiency of the drug is over 30% lower than that of the corresponding lead. Structures, targets and binding data for these pairs are reported in Table 5. Interestingly, in 3 of the 6 cases the target is a protein kinase. The binding affinities of lead and drug are very similar in 4 of these pairs, while the drug is considerably more potent than the lead in the other 2. A simple analysis of the 2D structures reveals a common pattern within this subset: each of the drugs has at least one additional charged group relative to the corresponding lead. This common feature suggests that these molecules were functionalized or further substituted in the late stages of the optimization process to improve some key properties without affecting binding to the target. Charged groups are commonly added to enhance solubility, and it is not surprising that such a measure would be required in the optimization of kinase inhibitors, which tend to be lipophilic and highly insoluble. The target-bound crystal structures are available for 4 of the 6 drugs in this subset, and snapshots from each of those structures are presented in Figure 10. A quick inspection of these structures reveals that the additional charged group makes limited or no interactions with the target in all four cases, thus providing no contribution to the binding energy and consequently reducing binding efficiency. In the case of lapatinib the side chain containing the charged group is so floppy and unconstrained by interactions with the target that its position cannot be defined crystallographically, thus resulting in a "truncated" crystal structure. The accounts detailing the discovery of these four drugs confirm that the charged moieties were incorporated to address property-related issues ranging from limited solubility to high protein binding or insufficient cellular concentration.[23−26] In the case of alvimopan, one of the two drugs for which the target-bound structure is not available, the authors report that the addition of a charged group was part of a strategy to minimize blood−brain barrier penetration.[27] Overall the analysis of this subset highlights the compromises that are often made at the late stages of drug discovery programs, when binding efficiency may have to be sacrificed in order to modulate specific properties without affecting potency.

**Lessons Learned and Suggested Guidelines for Lead Selection and Optimization.** The key findings of this study can be summarized as follows:

(1) On average, drugs and corresponding leads have similar binding efficiencies but significant increases or decreases along the lead optimization path are common.

(2) 90% of leads have binding efficiencies over 12.4, and 90% of drugs have binding efficiencies over 14.7.

(3) Leads with efficiencies as low as 6.8 have been shown to be viable if there is a clear design rationale.

(4) On average, $\mathrm{p}K_\mathrm{i}(\mathrm{drug}) \gg \mathrm{p}K_\mathrm{i}(\mathrm{lead})$ but $\mathrm{ClogP}(\mathrm{drug}) = \mathrm{ClogP}(\mathrm{lead})$, resulting in $\mathrm{LLE}(\mathrm{drug}) \gg \mathrm{LLE}(\mathrm{lead})$: a significant increase in lipophilic ligand efficiency is one of the recurring trends of successful drug discovery programs.

(5) Increasing molecular weight to improve potency is often inevitable. Maintaining similar lipophilicity when increasing size is one of the keys to the success of lead optimization programs.

(6) A large increase in binding efficiency (30% or more) can be achieved even when the lead scaffold is retained.

(7) Knowledge of the 3D structure of the target can help identify the hot spots in the active site where binding efficiency can be increased, thus lowering the requirement for an efficient starting point.

(8) Charge−charge interactions are often responsible for large efficiency boosts, but other factors can also contribute.

(9) Dissecting inefficient leads to then build on the most efficient fragments can be an effective strategy at the early stages of lead optimization.

(10) Binding efficiency can be effectively traded to achieve improved properties (e.g., higher solubility, reduced protein binding) in the late stages of a lead optimization program.

## Conclusions

The availability of a wealth of information on the structures and properties of known drugs enabled the derivation of solid and well-established rules defining druglikeness.[28,29] However, the understanding of what makes a good drug lead is not equally advanced, partly due to the lack of a broad and validated data set of genuine drug leads and their properties. Oprea and colleagues presented the first organized data set of lead/drug pairs and highlighted some of the key differences and similarities between leads and drugs.[2,30] The present study expanded on the comparative analysis of leads and drugs by incorporating binding data and focusing on binding efficiency as one of the key parameters. Previous studies on the binding efficiency trends in medicinal chemistry and drug discovery programs suggested that the efficiency of an optimal and minimally substituted lead scaffold can be retained at best and will likely decrease in the course of lead optimization. The present analysis produced interesting and sometimes surprising findings. A number of cases were identified in which the efficiency of the lead was largely improved while retaining the original scaffold and increasing the size, thus showing that the efficiency of an optimal core scaffold does not constitute a ceiling for a given congeneric series. Contrary to commonly accepted dogma, it was also shown that on average drugs and their originating leads have similar lipophilicity in spite of significant differences in size and potency, thus showing that increasing lipophilic binding efficiency can be one of the keys to a successful lead optimization process. Tentative thresholds for the acceptable binding efficiencies of putative leads and prospective drugs emerged from the analysis, which also highlighted the importance of dissecting the lead structure to build on the most efficient fragments when highly efficient starting points cannot be immediately identified.

Overall this study represents another step toward a better understanding of what makes a good lead structure and what are the most effective strategies to optimize that structure. Continued mining of the literature and a thorough documentation of ongoing drug discovery efforts should be encouraged so that broader data sets can be generated and the trends and guidelines presented here can be validated, reassessed or further refined.

**Acknowledgment.** I thank Dr. W. Patrick Walters for valuable advice on key aspects of the study and help with the software used in the analysis. I also thank Dr. Paul S. Charifson for careful reading of the manuscript and helpful suggestions.

**Supporting Information Available:** Excel and pdf files containing structures, affinity data, and corresponding references for the leads and drugs used in the study; text file containing 2D structures in SD format. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9997–10002.

(2) Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T. The design of leadlike combinatorial libraries. *Angew. Chem., Int. Ed.* **1999**, *38*, 3743–3748.

(3) Navia, M. A.; Chaturvedi, P. R. Design principles for orally bioavailable drugs. *Drug Discovery Today* **1996**, *1*, 179–189.

(4) Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235–249.

(5) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **2004**, *9*, 430–431.

(6) Abad-Zapatero, C.; Metz, J. T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discovery Today* **2005**, *10*, 464–469.

(7) Hajduk, P. J. Fragment-based drug design: how big is too big? *J. Med. Chem.* **2006**, *49*, 6972–6976.

(8) Reynolds, C. H.; Bembenek, S. D.; Tounge, B. A. The role of molecular size in ligand efficiency. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4258–4261.

(9) Hann, M. M.; Leach, A. R.; Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856–864.

(10) Murray, C. W.; Verdonk, M. L. The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 741–753.

(11) *Marvin*, version 5.0; ChemAxon: Budapest, Hungary, 2009.

(12) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific LLC: Palo Alto, CA, 2009.

(13) http://www.centerwatch.com/drug-information/fda-approvals.

(14) Das, J.; Chen, P.; Norris, D.; Padmanabha, R.; Lin, J.; Moquin, R. V.; Shen, Z.; Cook, L. S.; Doweyko, A. M.; Pitt, S.; Pang, S.; Shen, D. R.; Fang, Q.; de Fex, H. F.; McIntyre, K. W.; Shuster, D. J.; Gillooly, K. M.; Behnia, K.; Schieven, G. L.; Wityak, J.; Barrish, J. C. 2-Aminothiazole as a novel kinase inhibitor template. Structure−activity relationship studies toward the discovery of *N*-(2-chloro-6-methylphenyl)-2-[[6-[4-(2-hydroxyethyl)-1-piperazinyl)]-2-methyl-4-pyrimidinyl]amino)]-1,3-thiazole-5-carboxamide (dasatinib, BMS-354825) as a potent pan-Src kinase inhibitor. *J. Med. Chem.* **2006**, *49*, 6819–6832.

(15) Cockerill, S.; Stubberfield, C.; Stables, J.; Carter, M.; Guntrip, S.; Smith, K.; McKeown, S.; Shaw, R.; Topley, P.; Thomsen, L.; Affleck, K.; Jowett, A.; Hayes, D.; Willson, M.; Woollard, P.; Spalding, D. Indazolylamino quinazolines and pyridopyrimidines as inhibitors of the EGFr and C-erbB-2. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1401–1405.

(16) Zimmermann, J.; Caravatti, G.; Mett, H.; Meyer, T.; Muller, M.; Lydon, N. B.; Fabbro, D. Phenylamino-pyrimidine (PAP) derivatives: a new class of potent and selective inhibitors of protein kinase C (PKC). *Arch. Pharm.* (*Weinheim, Ger.*) **1996**, *329*, 371–376.

(17) Zimmermann, J.; Buchdunger, E.; Mett, H.; Meyer, T.; Lydon, N. B.; Traxler, P. Phenylamino-pyrimidine (PAP) derivatives: a new class of potent and highly selective PDGF-receptor autophosphorylation inhibitors. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1221–1226.

(18) Zimmermann, J.; Buchdunger, E.; Mett, H.; Meyer, T.; Lydon, N. B. Potent and selective inhibitors of the Abl-kinase: phenylamino-pyrimidine (PAP) derivatives. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 187–192.

(19) Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discovery* **2007**, *6*, 881–890.

(20) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.

(21) von Itzstein, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Van Phan, T.; Smythe, M. L.; White, H. F.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **1993**, *363*, 418–423.

(22) Kim, C. U.; Lew, W.; Williams, M. A.; Liu, H.; Zhang, L.; Swaminathan, S.; Bischofberger, N.; Chen, M. S.; Mendel, D. B.; Tai, C. Y.; Laver, W. G.; Stevens, R. C. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.* **1997**, *119*, 681–690.

(23) Barker, A. J.; Gibson, K. H.; Grundy, W.; Godfrey, A. A.; Barlow, J. J.; Healy, M. P.; Woodburn, J. R.; Ashton, S. E.; Curry, B. J.; Scarlett, L.; Henthorn, L.; Richards, L. Studies leading to the identification of ZD1839 (IRESSA): an orally active, selective epidermal growth factor receptor tyrosine kinase inhibitor targeted to the treatment of cancer. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1911–1914.

(24) Lackey, K. E. Lessons from the drug discovery of lapatinib, a dual ErbB1/2 tyrosine kinase inhibitor. *Curr. Top. Med. Chem.* **2006**, *6*, 435–460.

(25) Sun, L.; Liang, C.; Shirazian, S.; Zhou, Y.; Miller, T.; Cui, J.; Fukuda, J. Y.; Chu, J. Y.; Nematalla, A.; Wang, X.; Chen, H.; Sistla, A.; Luu, T. C.; Tang, F.; Wei, J.; Tang, C. Discovery of 5-[5-fluoro-2-oxo-1,2- dihydroindol-(3*Z*)-ylidenemethyl]-2,4-dimethyl-1*H*-pyrrole-3-carboxylic acid (2-diethylaminoethyl)amide, a novel tyrosine kinase inhibitor targeting vascular endothelial and platelet-derived growth factor receptor tyrosine kinase. *J. Med. Chem.* **2003**, *46*, 1116–1119.

(26) Kingsbury, W. D.; Boehm, J. C.; Jakas, D. R.; Holden, K. G.; Hecht, S. M.; Gallagher, G.; Caranfa, M. J.; McCabe, F. L.; Faucette, L. F.; Johnson, R. K.; Hertzberg, R. P. Synthesis of water-soluble (aminoalkyl)camptothecin analogues: inhibition of topoisomerase I and antitumor activity. *J. Med. Chem.* **1991**, *34*, 98–107.

(27) Zimmerman, D. M.; Gidda, J. S.; Cantrell, B. E.; Schoepp, D. D.; Johnson, B. G.; Leander, J. D. Discovery of a potent, peripherally selective *trans*-3,4-dimethyl-4-(3-hydroxyphenyl)piperidine opioid antagonist for the treatment of gastrointestinal motility disorders. *J. Med. Chem.* **1994**, *37*, 2262–2265.

(28) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25.

(29) Egan, W. J.; Merz, K. M., Jr.; Baldwin, J. J. Prediction of drug absorption using multivariate statistics. *J. Med. Chem.* **2000**, *43*, 3867–3877.

(30) Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308–1315.